Temporal dynamics of early visual word processing - early versus late N1 sensitivity in children and adults[I]

Aleksandra K. Eberhard-Moscicka[1,2], Lea B. Jost[1,2], Lynn V. Fehlbaum[1], Simone E. Pfenninger[3], & Urs Maurer [1,4,5#]

[1]Department of Psychology, Division of Cognitive Neuroscience, University of Zurich, Switzerland
[2]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
[3]Department of English and American Studies, University of Salzburg, Austria
[4]Department of Psychology, The Chinese University of Hong Kong, Hong Kong
[5]Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong

[#] U. M. is currently working at the Chinese University of Hong Kong

Abstract

In the course of reading development children become familiar with letter strings and learn to distinguish between lexical and non-lexical items. In previous studies, the N1 component of the ERP was shown to reflect print tuning but also to be sensitive to lexical effects. It remains unclear, however, whether these two aspects of orthographic processing occur at the same time or in different time windows during the lengthy N1 component. Moreover, it is unclear whether these processes develop late or occur already at early stages of literacy acquisition and whether this is similar for native languages and languages acquired later in life. To address these questions, 27 children were tested longitudinally, i.e. before (mean: 7.6 years) and after one year of classroom-based English instruction. Additionally, 22 adult speakers of English as a foreign language (mean: 25.1 years) were investigated. A 128-channel EEG was recorded while participants performed a one-back task with native German words, English words, pseudowords and false-font strings. The event-related EEG analysis of early and late N1 phases revealed early effects related to print tuning and late effects related to lexical processing in the native, but not in the second language of adult readers. In the absence of lexicality effects in

[I]Eberhard-Moscicka et al. (2016). *Neuropsychologia*, 91, 501-518.
doi: 10.1016/j.neuropsychologia.2016.09.014.

children, print tuning effects were found across both early and late N1 segments. The temporally distinct N1 sensitivities to print and lexicality reflect temporal dynamics of visual word processing, which seem to depend on reading expertise or maturation.

1. Introduction

From an early age we are constantly exposed to written words. With regard to progressing reading abilities, the vast majority of individuals become adept at effortless perception and decoding of visual word forms. Skilled readers can process complicated stimuli and distinguish between orthographic and non-orthographic stimuli in order to promptly access their linguistic representations. This acquired expertise underlies rapid and efficient identification of letters and words within visual networks of our native language and, presumably, of additional languages learned later in life.

1.1. N1 visual word processing

Recent years have provided an extensive body of research delivering evidence that reading abilities are supported by a highly specialized functional organization that develops early in the course of reading training. A cortical region in the left occipito-temporal cortex that is not sensitive to low-level visual word properties such as size, position, font, or letter case (Cohen et al., 2000; McCandliss et al., 2003) was proposed to be a center of visual word form recognition (Cohen & Dehaene, 2004). This visual word form area (VWFA) was not considered to represent an 'in-built' module but an outcome of progressive specialization for visual word recognition (Cohen & Dehaene, 2004), which is being activated whenever literate subjects are presented with letter strings (Aghababian & Nazir, 2000; Brem et al., 2009; Maurer et al., 2005a; Maurer & McCandliss, 2007; McCandliss et al., 2003), specifically in a familiar orthography (Baker et al., 2007).
Converging functional neuroimaging and electrophysiological evidence has demonstrated that this selectivity of the VWFA originates from extensive experience with visual word forms. The N1 (or N170) component of the event-related potential (ERP) has been reported to

show prelexical sensitivity to orthographic versus non-orthographic strings (Bentin et al., 1999; Gros et al., 2002; Maurer et al., 2005a; Rossion et al., 2003; Simon et al., 2004), presumably linked to the VWFA (Brem et al., 2006). While being absent in non-reading kindergarten children (Maurer et al., 2005a), this sensitivity emerged after short grapheme-phoneme training even in kindergartners with no prior reading experience (Brem et al., 2010), but was consolidated only after one year of formalized classroom-based reading instruction (Eberhard-Moscicka et al., 2015). This neurophysiological sensitivity, manifested in adults as negative deflection typically over left occipito-temporal channels at about 170 ms, is thus believed to reflect fast perceptual specialization for print, which is also referred to as coarse print tuning (e.g., Brem et al., 2006; Eberhard-Moscicka et al., 2015; Maurer et al., 2006; Maurer et al., 2011).

Developmentally, the initially large N1 amplitudes and longer latencies to both orthographic and non-orthographic stimuli in the early stages of the acquisition of reading skills attenuate in older, more experienced readers (Brem et al., 2009; Maurer et al., 2006).

Notably, both electroencephalographic (EEG) and magnetoencephalographic (MEG) studies that used symbol strings as a control condition consistently reported robust print tuning effects with words eliciting more activation than symbol strings, irrespective of the population's age (e.g., Bentin et al., 1999; Brem et al., 2006; Brem et al., 2009; Brem et al., 2005; Eberhard-Moscicka et al., 2015; Maurer et al., 2005a; Maurer et al., 2006; Maurer et al., 2011; Parviainen et al., 2006; Tarkiainen et al., 1999). However, studies using visually more closely matched false-font characters are still scarce and the respective findings vague. While false-font studies with adult participants do not always reveal robust print tuning effects (Eulitz et al., 2000; Schendan et al., 1998; Wong et al., 2005; Xue et al., 2008), studies with children consistently indicate stronger activation in response to words than to false-font strings (Brem et al., 2010; Hasko et al., 2013). As such, it appears imperative at this point to investigate the developmental course of print tuning using false-font characters, as children become more skilled readers.

Additionally, the lexical contrast between words and pronounceable non-words (hereafter pseudowords) reveals weaker and less robust effects in the N1 component, suggesting that task demands (be they implicit: Maurer at al., 2005b; Proverbio et al., 2009; or explicit: Hauk et al., 2006b; Hauk et al., 2006a), as well as additional factors such as language and development (Maurer et al., 2006) might influence the lexical effects within the N1 component. Accordingly, N1 lexicality effects may be more prominent at the initial stages of reading acquisition, similar to the inverted u-curve of the N1 print tuning development (Brem et al., 2009; Maurer et al., 2006).

Importantly, there is evidence that several processes coincide within the N1 component in response to visual words (Korinth et al., 2011, 2012; Simon et al., 2004). This is also suggested by different lateralization patterns between the early and late parts of the N1 (Appelbaum et al., 2009; Cohen et al., 2000; Maurer et al., 2005a; Nemrodov et al., 2011) and the observation of distinct peaks at 160 and 210 ms with posterior negativity (Hauk et al., 2006b), which typically coincides with the broad N1 time window in previous studies (Brem et al., 2009; Mahé et al., 2012; Maurer et al., 2005b; Maurer et al., 2005a). Hence, dividing the usually lengthy N1 time segment into two parts could provide more comprehensive understanding of the temporal dynamics of early visual word processing and increased sensitivity to familiarity of word forms, as indicated by differences between words and pseudowords.

An area where initially unfamiliar word forms become familiar is the area of foreign language learning. Thus, longitudinal studies on foreign language acquisition could be revealing not only in terms of language-related processes but also in terms of lexical processing, as the same stimuli are once perceived as pseudowords (before foreign language training) and once as words (after foreign language training).

1.2. Foreign language learning and N1 visual word processing

To date, several studies have discussed the effects of foreign language learning and phoneme processing (e.g., Jost et al., 2015; Näätänen et al., 1997; Rinker et al., 2010; Shestakova et al., 2003; Winkler et al., 1999). However, so far not much research has been addressed to the issue of foreign language learning and visual word processing. While a few studies investigated late components (i.e., N400 and P600; see McLaughlin et al., 2004; Osterhout et al., 2008; Stein et al., 2006), evidence of foreign language acquisition and the N1 effects is still scarce and primarily focused on the context of short-term artificial language training and development of print tuning (Maurer et al., 2010).

Whereas few studies investigated lateralization of neural circuits involved in written word recognition for the first and second language (Grossi et al., 2010; Proverbio et al., 2002), more subtle between-language contrasts were addressed by Proverbio et al. (2009). In this study, the timing of brain activation during linguistic processing of native versus later-acquired languages in simultaneous interpreters of three languages sharing the same alphabet was investigated. Although the subjects were native-like speakers of additional languages they had acquired later in life, the occipito-temporal N1 effects were found for the native language only. Moreover, at around 200 ms, the parietal N1 differed between words of the three languages, showing a proficiency-related amplitude increase (i.e. largest N1 amplitudes for the native language). Hence, the N1 seems sensitive not only to lexical differences but also to effects of proficiency following foreign language learning. As these effects were rather small, it is plausible that the lexical effects are larger at the early stages of learning a foreign language and attenuate with progressive proficiency.

The goal of the present study was to take this discussion a step further by investigating processing of known vs. newly learned visual word forms. Moreover, we aimed at testing the print tuning effects with the use of false-font strings, which, in contrast to symbol strings, constitute a better control for low-level visual differences. To this end, we assessed 27 native (Swiss-) German primary school children in a longitudinal fashion, i.e. before they began formal, classroom-based

English instruction (1st grade) and one year into learning English as a foreign language (3rd grade). Additionally, we investigated 22 native (Swiss-) German adult speakers of English as a foreign language. We focused on early and late parts of the N1 component of the ERP that has been shown to play an important role in visual word processing. The specific goals of the study are two-fold: (a) to assess the temporal dynamics of print tuning in children in a longitudinal fashion and investigate the early effects of classroom-based English instruction on visual word processing within the N1 electrophysiological response, and (b) to recognize the pattern of temporal dynamics of the N1 print tuning and lexicality effects in adult speakers of English as a foreign language. As ERPs reveal stimulus-related activity in the cortex in millisecond resolution, they provide an excellent tool for studying the temporal dynamics of functional selectivity related to early linguistic processes in the brain.

2. Methods

2.1. Participants

We report data of 27 native (Swiss-) German-speaking children (12 girls and 15 boys; 3 left-handed). Children were tested longitudinally; the first assessment took place prior to formal English instruction in school (i.e. at the end of 1st grade, mean ± *SD*, years = 7.55 ± 0.31), whereas the second assessment took place after one year of English instruction (i.e. at the beginning of 3rd grade, mean ± *SD*, years = 8.88 ± 0.33). This way we assured that every child went through a full year of classroom-based English instruction (mean time elapsed between first and second assessment was 478 days (*SD*=22), i.e. 1 year and 3 months on average). From an original group of 51 monolingual children, one dropped out of the study, one transferred to a bilingual school, five had to be excluded due to dyslexia (see criteria below) and 17 were excluded due to low number of accepted trials in the English condition (below 26 trials i.e. 30% of the total number of 84 trails, see section 2.3.2). None of the remaining subjects needed to be excluded due to low signal-to-noise ratio (i.e. none of the children was below the threshold of 2 for any of the four conditions (P1/N1 components)).

Moreover, we assessed 22 native (Swiss-) German adult (mean ± *SD*, years = 25.14 ± 3.67; 13 females and 9 males; 3 left-handed) speakers of English as a foreign language (their English vocabulary size assessed as correctly scored high-frequency words in percentage of total items: mean ± *SD*, 81.6%, ± 15.5; Productive Vocabulary Levels Test, PVLT, Laufer & Nation, 1999). All subjects had normal or corrected-to-normal vision, and every child had an estimated non-verbal IQ equal or above 80, i.e. not more than 1.333 *SD* below the normative mean in HAWIK-IV (*M*=100, *SD*=15, subtest: block design, Petermann & Petermann, 2010, corresponding to the English version of the Wechsler Intelligence Scale for Children). The adult participants scored above the 10^{th} percentile of the norms in a standardized reading test (SLRT-II, Moll & Landerl, 2010). For the children, unimpaired reading was assumed if their reading fluency was above the 10^{th} percentile in standardized reading-fluency tests (SLS 1-4, Mayringer & Wimmer, 2005; SLRT-II, Moll & Landerl, 2010). The study protocol was in agreement with the local ethics committee. Consent was obtained orally from children and in written form from their parents and adult participants. Moreover, children's parents and adult participants filled out a background questionnaire screening for a history of neurological diseases and psychiatric disorders.

2.2. Procedure

All subjects participated in a behavioral session and an EEG session. While adults were tested once, children were tested longitudinally, i.e. prior to formal classroom-based English instruction (1^{st} grade) and one year thereafter (3^{rd} grade, see section 2.1.). The behavioral session consisting of a set of language tasks in the native and English language (which is beyond the scope of this paper, hence, shall not be discussed in further detail) took approximately one hour for the adult participants and about one and a half hours for first and third grade children. The EEG assessment, which was identical for all the groups at each testing time (see section 2.3.), was administered using one of two portable EEG systems (Electrical Geodesics, Inc, EGI). The EEG session was approximately three and a half hours long and took place either in the

EEG laboratory at the Department of Psychology at the University of Zurich or in a separate room provided by the schools. Before using a room at schools, a standard quality check was made in order to ensure the absence of 50 Hz noise. As compensation for their participation, children received a written report about their reading skills as well as a book voucher of 20 CHF at each testing time (i.e. in 1st and 3rd grade after both behavioral and EEG sessions). Adults either received course credits or 50 CHF for their participation in the study.

2.3. EEG recording

The EEG session was identical for first and third grade children as well as for adult participants. Prior to the main EEG recording, subjects were seated 80 cm away from the computer screen. They were instructed on task demands and performed a practice experimental run that lasted about 1 min. A one-back task (see Fig. 1, approx. 20mins. long) assessing visual processing of alphabetic strings was part of a larger session that included several experiments presented in a pseudo-randomized order. Participants were allowed to take extensive breaks between experiments, which resulted in a session duration of about 3.5 hours. During experiments, compliance was monitored with a digital camera.

2.3.1. Material and stimuli

The stimuli presented were familiar German words (high frequency of occurrence according to ChildLex – a text corpus encompassing 6-8 year old children's print language in German: *M*=161.86/Mio, ChildLex Lexical Database, Schroeder et al., 2015; and according to CELEX: *M*=233.78/Mio, CELEX Lexical Database, Baayen et al., 1993; implemented in WordGen application, Duyck et al., 2004), familiar English words (high frequency of occurrence according to CELEX: *M*=221.21/Mio, CELEX Lexical Database, Baayen et al., 1993; implemented in WordGen application, Duyck et al., 2004; German and English words were matched on lexical frequency *p*=.90), pseudowords (created from letters that appeared in German and English stimuli; pronounceable in German and English with higher orthographic neighbourhood size in English than German (*M*: 7.4 vs. 3, *p*=.024), CELEX Lexical Database, Baayen et al., 1993; implemented in WordGen application, Duyck et al., 2004), and false-font strings (matched with German words). The false-font characters, where each alphabetic letter had its unique false-font correspondent were designed for the purpose of this study. Visual complexity was controlled for by keeping the constant number of lines (e.g.: **N** = $\barwedge$) or by keeping the overall configuration of the alphabetical counterpart (e.g.: **U** = $\cap$) . In order to ensure that after one year of formal classroom-based English instruction (i.e. in 3rd grade) children would have learned the English words presented, we selected them from the current compulsory English teaching materials for second grade pupils in the Canton of Zurich (Interkantonale Lehrmittelzentrale, 2005). Due to the limited number of English words that we expected children to know at the second data collection time in third grade and due to additional matching constraints (see below), we limited the number of items per condition to 14. The 14 stimuli per condition (see S1) were repeated six times (84 stimuli per condition) and presented in six blocks (the order of conditions was counterbalanced). Additionally, 12 immediate repetitions serving as targets were presented in each condition. The 12 targets per condition (i.e. German words, English words, pseudowords and false-font strings) were pseudo-randomly selected from the 14 stimuli in a way that no stimulus was repeated more than once. Moreover, to reduce participant's expectancy, the number of repetitions per condition and per block was randomized (i.e. in a single condition block the number of repetitions varied between 1 to 3). All experimental characters were presented in black (Arial, bold, font size 28) and occurred in the center of a white rectangular box (85 mm x 47 mm) in the middle of a grey background. Each stimulus appeared for 500 ms and was followed by a mean interstimulus interval of 1500 ms (jittered between 1250-1750 ms). Across the conditions, the stimuli were matched for string length and contained 3.9 letters/false-font characters on average (range: 3-5; average length and height: 31.9 mm x 7 mm). In addition, German and English words as well as pseudowords were matched for number of letters, frequency of letters and number of syllables. Moreover, according to a text corpus encompassing 6-8-year-old children's print language in German (ChildLex Lexical Database, Schroeder et al., 2015), bigram frequency was not significantly different either for a comparison of German words (*M*=31965.36) versus pseudowords (*M*=25980.89) (*p*=.304), or for German words versus English words (*M*=30148.93) (*p*=.725), or for English words versus pseudowords (*p*=.453).

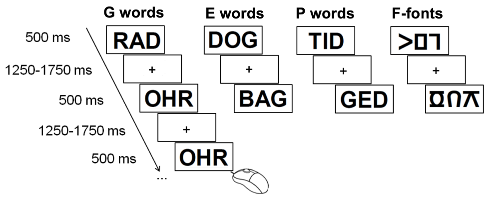


Fig.1. In a one-back task participants were instructed to watch the stimuli and press the mouse button for immediate repetitions (12.5% immediate repetitions presented randomly in each stimulus condition). Stimuli were presented for 500 ms and were followed by a mean interstimulus interval of 1500 ms (jittered between 1250-1750 ms). G words = German words; E words = English words; P words = pseudowords; F-fonts = false-font strings.

2.3.2. Electrophysiological recording and analysis

A 128-channel EEG (HydroCel GSN, EGI NA 300 amplifier) was recorded against the Cz reference, at a sampling rate of 250 Hz, with high- (0.1 Hz) and low-pass (100 Hz) filter settings. As modern high-input impedance amplifiers and their accurate digital filters for power noise provide excellent EEG signal collection even with higher electrode impedances (Ferree et al., 2001), the electrode impedance was kept below 50 kΩ. This impedance threshold has been widely used in previous studies using EGI systems across a variety of EEG labs (e.g., Franklin et al., 2007; Hämäläinen et al., 2015; Maurer et al., 2005b; Rihs et al., 2007). Raw data was preprocessed using BESA software. The continuous EEG was corrected for eye blinks after channels with extensive artifacts were spline interpolated. Corrected files were digitally low-pass (30 Hz) and high-pass filtered (0.3 Hz) and segmented (-150 ms prior and 850 ms following the stimulus). At the end of the EEG session every participant was asked to read the list of English words that were presented in the one-back experiment and indicate the meaning of each word. English words that were familiar to first graders were excluded from further analysis on an individual basis. For third grade children and for adult participants, only the known English words were included in the analysis. This procedure allowed us to ensure that the event-related potentials to English words that were included in the final analysis were the ERPs evoked by English words that were familiar to third grade children and adults, while being unfamiliar to first grade children. Furthermore, trials with artifacts exceeding the max-min difference of 180 μV (for children) and 120 μV (for adults) in any channel were automatically excluded before averaging. Moreover, only participants whose accepted trial ratio was above 30% (i.e. at least 26 trials in each of the four conditions) were included in the analysis. Accordingly, the mean number of accepted trials (with standard deviations in parentheses) in the grand averages (per group and condition) were as follows: children in 1st grade: German words 59.56 (13.82), pseudowords 59.78 (13.32), English words 59.04 (13.67) and false-font strings 61.44 (11.42); children in 3rd grade: German words 65.41 (10.11), pseudowords 64.37 (10.81), English words 39.07 (9.96) and false-font strings 65.67 (7.97); and adults German words 75.36 (9.30), pseudowords 78.05 (5.79), English words 74.41 (8.19) and false-font strings 78.82 (5.10). The lower number of trials in the English word condition in 3rd graders was due to the exclusion of words that the children did not know in the vocabulary knowledge test that followed the EEG session (see above). Nevertheless, despite the lower number of trials, the ERP waveforms of this condition were remarkably similar to the waveforms in the German word and pseudoword conditions.

The data was later transformed to the average reference (Lehmann & Skrandies, 1980) and the recording reference was used as an additional electrode for further data processing. Similar to Pegado et al. (2014), the ERPs were corrected for the amplifier delay of 8 ms (induced by the anti-aliasing filters of EGI NA 300 amplifiers with the current sampling rate, Advisory Notice, 29 August 2014, Electrical Geodesics Inc.) and a constant 20 ms delay (as revealed by a timing test using a photo sensor). After this procedure, the component latencies were within expected time ranges (i.e. P1 in adults ~100 ms; N1 in adults ~160 ms). In the final pre-processing step, the ERPs of all four conditions (i.e. German words, English words, pseudowords and false-font strings) were averaged separately, after the target stimuli from the one-back task were automatically excluded.

Using Brain Vision Analyzer Software, the individual ERPs were baseline-corrected, and Global Field Power (GFP) together with grand means for all the four condition stimuli (i.e. German words, English words, pseudowords and false-font strings) were computed separately for each experimental group (i.e. 1st and 3rd grade children as well as adults).

We investigated temporal unfolding of early visual word processing for print tuning and lexicality effects across the N1 segment. The N1 segment was defined for each group separately and based on two GFP minima of the grand mean (averaged over: German words and false-font strings for print tuning; German words, English words and pseudowords for lexicality effects; see also Maurer et al., 2005a). The GFP minima reflect the start and the end points of ERP components that occur at the level of entire maps (i.e., across all electrodes). This is a common approach for choosing time windows in an unbiased way (e.g. Albrecht et al., 2005; Meyer et al., 2007; Maurer et al., 2005a). Time windows for the print tuning and the lexicality effects were determined separately, to ensure that the conditions to be compared contributed equally to the selection. As such, print tuning was indexed by the N1 difference in the ERPs between German words and false-font strings, while the lexicality effect in the native language (L1) was indexed by the N1 difference in the ERPs between German words and pseudowords, and the lexicality effect in the second language (L2) was indexed by the N1 difference in the ERPs between English words and pseudowords. In order to gain a more comprehensive understanding of temporal dynamics of visual word processing across the entire N1 segment, the N1 was subdivided into two parts (i.e. early and late N1; for detailed information on time segments see Fig. 2), across which the voltage values were averaged (see also Brem et al., 2006; Cohen et al., 2000; Nemrodov et al., 2011). The early and the late N1 time segments were defined by dividing the N1 time window into two segments of equal length. This choice was further supported by visual inspection of the ERP grand averages, a normalized TANOVA revealing distinct topographies between the early and late N1 time windows in adults (German words: p=.061; false-font strings: p=.004; English words: p=.019; pseudowords: p=.095) and by sample-by-sample t-tests (for details on the sample-by-sample t-tests see panel C. Fig. 2).

The analyses for print tuning as well as lexicality effects in L1 and L2 were performed on the left occipito-temporal electrode cluster (LOT: E50, E57, E58, E59, E63, E64, E65, E66, E68, E69, E70, E73, E74; see Fig. 2). Corresponding electrodes in the 10-10 system (Luu & Ferree, 2000) are P7, P9, TP9, PO7, PO9, O1. The choice of the left occipito-temporal electrode cluster relies upon previous investigations indicating higher left-hemispheric sensitivity to print (e.g. Bentin et al., 1999; Brem et al., 2005; Eberhard-Moscicka et al., 2015; Parviainen et al., 2006; Tarkiainen et al., 1999). The analyses were performed on focal occipito-temporal electrodes rather than on entire ERP maps as previous studies indicated that the analyses on occipito-temporal electrodes (where the N1 showed the largest negativity, and thus the strongest signal) were more sensitive to effects of print tuning than the GFP measure (Maurer et al., 2007, 2011).

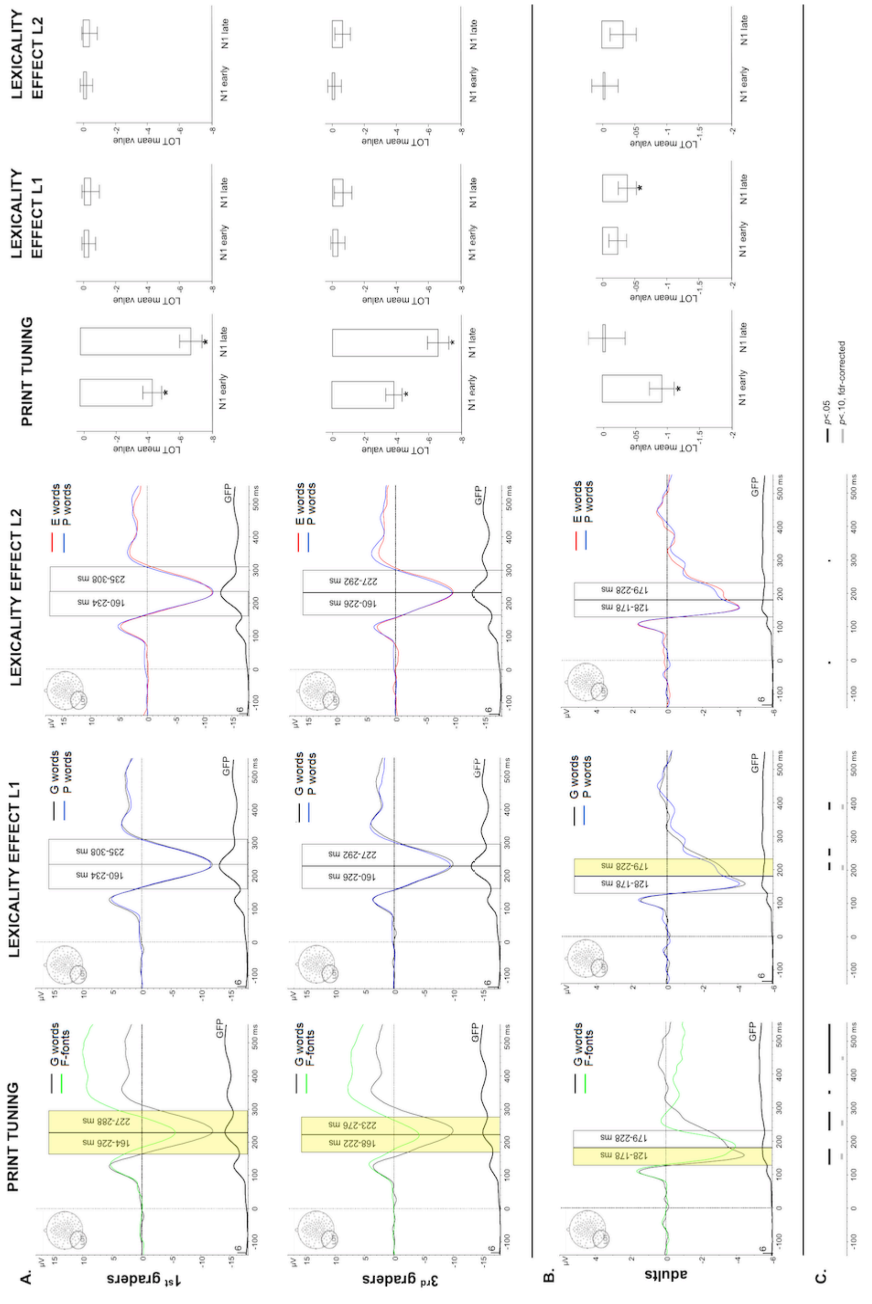

A.
PRINT TUNING
LEXICALITY EFFECT L1
LEXICALITY EFFECT L2
PRINT TUNING
LEXICALITY EFFECT L1
LEXICALITY EFFECT L2
1st graders
3rd graders
B.
adults
C.
G words
F-fonts
G words
P words
E words
P words
GFP
164-226 ms
227-288 ms
160-234 ms
235-308 ms
168-222 ms
223-276 ms
160-226 ms
227-292 ms
128-178 ms
179-228 ms
N1 early
N1 late
LOT mean value
p<.05
p<.10, fdr-corrected

Fig.7. ERP waveforms at left occipito-temporal electrode clusters (LOT) and bar graphs of the mean values at LOT with standard error bars (+/- 1 SE) for print tuning (German words vs. false-font strings), lexicality effect in the native language (German words vs. pseudowords) and lexicality effect in the L2 (English words vs. pseudowords) and the two N1 segments of interest. (A) First grade children (print tuning: early N1 164-226 ms, late N1 227-288 ms; lexicality effects: early N1 160-234 ms, late N1 235-308 ms), third grade children (print tuning: early N1 168-222 ms, late N1 223-276 ms; lexicality effects: early N1 160-226 ms, late N1 227-292 ms). (B) Adults (print tuning: early N1 128-178 ms, late N1 179-228 ms; lexicality effects: early N1 128-178 ms, late N1 179-228 ms). The significant time windows are indicated in yellow and asterisks depict significant post hoc planned comparisons at *$p<.0125$ Bonferroni corrected. (C) Results of the sample-by-sample t-tests for print tuning, lexicality effect L1 and lexicality effect L2 for adult participants. The black line indicates time windows at $p<.05$ (uncorrected), while the grey line indicates time windows corrected for multiple comparisons with the unified algorithm to false discovery rate (fdr) estimation (Strimmer, 2008; Yoncheva et al., 2013) at $p<.10$.

2.4. Statistical analysis

To investigate the temporal dynamics of print tuning and lexicality effects, an ANOVA with factors *segment* (early N1 vs. late N1) and *effect* (print tuning vs. lexicality L1) was computed for adult participants. The longitudinal effects of print and lexical processing in 1st and 3rd grade children were investigated with an ANOVA with within-subject factors *grade* (1st vs. 3rd), *segment* (early N1 vs. late N1) and *effect* (print tuning vs. lexicality L1). To answer the question whether lexical processing is any different in the native than in the L2, an ANOVA with factors *segment* (early N1 vs. late N1) and *language* (lexicality L1 vs. lexicality L2) was performed for adults, while an ANOVA with within-subject factors *grade* (1st vs. 3rd), *segment* (early N1 vs. late N1) and *language* (lexicality L1 vs. lexicality L2) was computed for children. As previous studies consistently reported more negative going amplitudes to word than to control stimuli (be they language or non-language stimuli) in a repetition detection task (e.g. Maurer et al., 2005; Maurer et al., 2006), planned comparisons with one-tailed t-tests were computed for print tuning and lexicality effects in L1 and L2. Bonferroni correction was applied for planned comparisons, and only the effects that remained significant after the Bonferroni correction are interpreted.

Difference t-maps for print tuning and lexicality effects in L1 and L2 were derived separately for each group and both early and late N1 segments (Fig. 3). For the difference t-maps, t-values of ± 2.06 in both groups of children and t-values of ± 2.08 in adult participants indicate a significant difference at $p<.05$. Moreover, topographic voltage maps of the grand-averaged ERPs to German words, English words, pseudowords and false-font strings were derived separately for each group and both early and late N1 segments to visualize how the scalp distribution changed over time (see supplementary Fig. S2).

For the analysis of the repetition-detection task, which was primarily applied to ascertain that participants paid attention to the stimuli presented, repeated measure ANOVAs were computed for each experimental group, separately for accuracy and reaction time with a within-subject factor *condition* (German words vs. false-font strings vs. English words vs. pseudowords). In order to gain a better understanding of the difference in performance between the condition pairs, post-hoc Bonferroni corrected t-tests were computed for the significant main condition effects.

3. Results

3.1. Longitudinal data of 1st and 3rd grade children

3.1.1. ERP data: Temporal dynamics of N1 print tuning and lexicality effects in children

We compared German word–false-font differences and German word–pseudoword differences in a repeated-measure ANOVA with the within-subject factors *effect* (print tuning vs. lexicality L1), *segment* (early N1 vs. late N1) and *grade* (1st vs 3rd grade). Print tuning was larger than L1 lexical effect (*effect,* $F(1,26)=151.32$, $p<.001$, $\eta_p^2=.85$), being more pronounced in the late compared to the early N1 (*segment* x *effect*, $F(1,26)=50.80$, $p<.001$, $\eta_p^2=.66$; see panel A. Fig. 2 and 8). Moreover, this interaction modulated the main effect of *segment* ($F(1,26)=39.13$, $p<.001$, $\eta_p^2=.60$). Neither the main effect of *grade* nor any other interactions were significant (all $F(1,26) \leq 0.81$, all $p \geq .378$ and all $\eta_p^2 \leq .03$).

Due to the absence of any statistically significant differences between 1st and 3rd grade children in the main ANOVA, the planned comparisons for the early and the late N1 print tuning and L1 lexicality effect were performed on the average values across 1st and 3rd grade children. A very robust print tuning effect was found in both early ($t(27)=-9.68$, $p<.001$, Bonferroni corrected $p<.0125$) and late N1 segments ($t(27)=-12.71$, $p<.001$, Bonferroni corrected $p<.0125$), while no L1 lexicality effect was found in either early ($t(27)=-0.98$, $p=.336$, Bonferroni corrected p=ns) or late ($t(27)=-1.22$, $p=.234$, Bonferroni corrected p=ns) N1 segment (see panel A. Fig. 3).

In order to investigate the learning effects of one year of formalized classroom-based English instruction, we compared L1 (German words vs. pseudowords) and L2 (English words vs. pseudowords) lexicality effects in a repeated-measure ANOVA with within-subject factors *language* (lexicality L1 vs. lexicality L2), *segment* (early N1 vs. late N1), and *grade* (1st vs. 3rd grade). Accordingly, no significant effects, and thus no lexicality-related learning effects were found (all $F(1,26) \leq 1.23$, all $p \geq .277$ and all $\eta_p^2 \leq .05$). Indeed, planned comparisons that were performed for 3rd grade children revealed no lexicality effects in the L2 in neither early nor late N1 segments (both $t(27) \leq -1.33$ and $p \geq .196$, Bonferroni corrected p=ns; see panel A. Fig. 3).

3.1.2. Behavioral data

The average accuracy and reaction time to target stimuli for German words, false-font strings, English words and pseudowords for 1st and 3rd grade children are reported in Table 1. According to repeated measure ANOVAs, 1st graders showed no significant differences in accuracy (*condition*, $F(1,26)=1.40$, $p=.267$, $\eta_p^2=.15$) and in reaction times (*condition*, $F(1,26)=1.85$, $p=.166$ $\eta_p^2=.19$). Interestingly, the same group of children still did not show reaction time differences in third grade (*condition*, $F(1,26)=1.71$, $p=.192$, $\eta_p^2=.18$), but they showed overall accuracy differences (*condition*, $F(1,26)=6.06$, $p=.003$, $\eta_p^2=.43$) with more accurate responses to German words ($t(1,26)=3.85$, $p=.001$, Bonferroni corrected *p*<.010) and English words (t(1,26)=3.85, *p*=.001, Bonferroni corrected *p*<.010) as compared to false-font strings. The remaining condition comparisons were not significant (all p>.514).

Table 1. Behavioral results for target stimuli for first and third grade children and adults.

	Condition	Accuracy (% ±*SD*)	Reaction time (ms ±*SD*)
1st graders	G words	73.3 (23.5)	967.3 (129.6)
	F-fonts	64.5 (23.4)	932.2 (129.4)
	E words	70.2 (29.2)	937.9 (229.7)
	P words	69.7 (23.2)	977.2 (145.1)
3rd graders	G words	87.1 (10.5)	836.5 (144.3)
	F-fonts	72.9 (20.1)	859.6 (178.7)
	E words	86.3 (16.2)	862.6 (141.2)
	P words	84.3 (19.1)	876.3 (150.2)
adults	G words	93.91 (8.30)	605.5 (113.03)
	F-fonts	91.6 (10.3)	590.8 (105.4)
	E words	95.83 (7.16)	610.0 (109.50)
	P words	97.73 (6.42)	630.9 (119.25)

G words = German words; F-fonts = false-font strings; E words = English words; P words = pseudowords

3.2. Data of adults

3.2.1. ERP data: Temporal dynamics of N1 print tuning and lexicality effects in adults

We compared German word–false-font differences and German word–pseudoword differences in a repeated-measure ANOVA with the within-subject factors *effect* (print tuning vs. lexicality L1) and *segment* (early N1 vs. late N1). Accordingly, N1 amplitudes were more negative in the early N1 segment compared to the late N1 segment (*segment,* F(1,21)=7.71, *p*=.011, η_p^2=.27). Importantly, while the print tuning effect was stronger in the early vs. late N1 segment, the L1 lexicality effect showed the reversed pattern, being more pronounced in the late rather than the early N1 segment (*segment* x *effect,* F(1,21)=12.99 *p*=.002, η_p^2=.38; see panel B. Fig. 2). Planned comparisons revealed significant print tuning effects in the early (t(22)=-4.82, *p*<.001, Bonferroni corrected *p*<.0125) but not in the late N1 (t(22)=-0.12, *p*=.904, Bonferroni corrected *p*=.ns). In contrast, L1 lexicality effects were significant in the late N1 ($t(22)=-2.72$, $p=.011$, Bonferroni corrected $p<.0125$), but not in the early N1 ($t(22)=-1.72$, $p=.101$, Bonferroni corrected p=.ns; see panel B. Fig. 3).

In order to explore whether lexical processing in the native language was any different to lexical processing in the second language we computed a repeated-measure ANOVA on word–pseudoword differences in L1 and L2 with within-subject factors *language* (lexicality L1 vs. lexicality L2) and *segment* (early N1 vs. late N1). Accordingly, we did not find any differences with respect to *segment* ($F(1,21)=3.36$, $p=.081$, $\eta_p^2=.14$) or *language* ($F(1,21)=0.49$, $p=.493$, $\eta_p^2=.02$), nor was their interaction significant (*segment* x *language,* $F(1,21)=0.79$, $p=.383$, $\eta_p^2=.04$). Indeed, contrary to the lexicality effects that were found for the late N1 segment in the L1 (see above), planned comparisons revealed no significant lexicality effects for L2 either in the early ($t(22)=-0.14$, $p=.893$, Bonferroni corrected p=ns) or late N1 ($t(22)=-1.64$, $p=.116$, Bonferroni corrected p=ns; see panel B. Fig. 3).

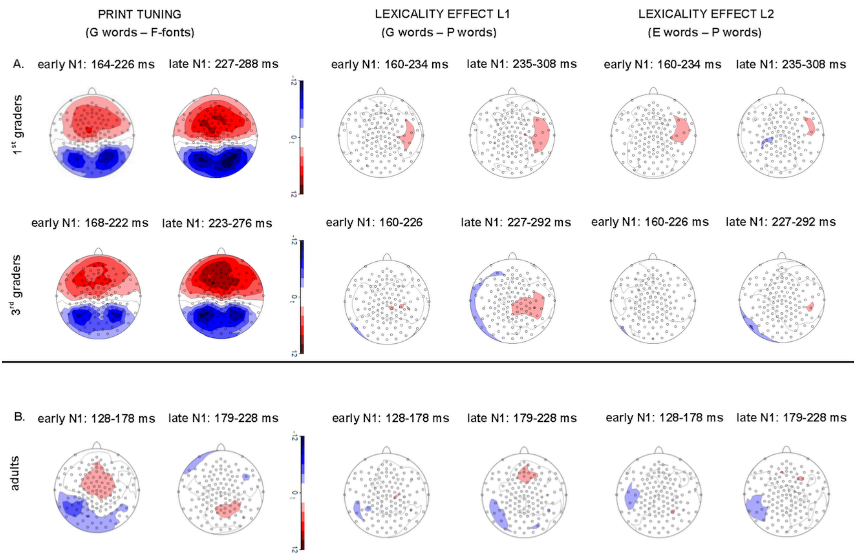


Fig. 3. Early and late N1 difference maps (t-maps) for print tuning (i.e. German words vs. false-font strings), lexicality effect in the native language (i.e. German words vs. pseudowords) and lexicality effect in L2 (i.e. English words vs.

pseudowords) for 1st and 3rd graders (A), and for adults (B). While the t-maps showed robust print-tuning effects with a typical N1 distribution in both segments of the children (t-values of ± 2.06 significant at $p<.05$) and in the early N1 of the adults (t-values of ± 2.08 significant at $p<.05$), lexicality effects with a typical N1 distribution (occipito-temporal negativity and fronto-central positivity) were only found in the late N1 in adults.

3.2.2. Behavioral data

The average accuracy and reaction time to target stimuli for German words, false-font strings, English words and pseudowords for adult participants are reported in Table 1. According to repeated measure ANOVAs, adults showed overall significant differences in accuracy (*condition*, $F(1,21)=3.69$, $p=.030$, $\eta_p^2=.37$), while showing no significant differences in reaction times (*condition*, $F(1,21)=1.08$, $p=.185$ $\eta_p^2=.27$). The follow-up t-tests on accuracy revealed that adults were more accurate in response to pseudowords compared to German words ($t(1,21)=-2.87$, $p=.009$, Bonferroni corrected $p<.010$), while the remaining comparisons were not significant (all p>.09).

4. Discussion

The present study aimed at investigating the temporal unfolding of the N1 visual word processing in native (Swiss)-German children tested longitudinally (i.e. prior to, and one year after, formal, classroom-based English instruction) and in native adult (Swiss)-German speakers of English as a foreign language. The longitudinal data of children indicated robust print tuning effects across both early and late N1 segments. Contrary to our expectations, young learners of English as a foreign language not only failed to show lexicality effects in their L2, they also did not display any lexicality effects in their native language, which could be indicative of language-related processes that have not been entirely established yet. While adult speakers of English as a foreign language exhibited print tuning in the early N1, lexicality effects were found in the late N1, indicating that sensitivity to print and lexicality unfolds differently throughout the N1 segment in adults. Although lexicality effects in adults did not differ between the native language and the English second language, lexicality effects in the L2 did not reach significance. This could suggest that lexical sensitivity requires a large amount of exposure and reading practice to develop in skilled readers.

4.1. Print tuning – robust effects in children and adults

Many previous studies on coarse print tuning used simple geometric forms as a control condition that were not well matched for low-level visual differences (e.g., Brem et al., 2006; Brem et al., 2009; Maurer et al., 2005b; Maurer et al., 2005a; Parviainen et al., 2006; Tarkiainen et al., 1999; Zhao et al., 2014). In studies using better matched false-font strings, print tuning was reliably found in children (Brem et al., 2010; Brem et al., 2013; Hasko et al., 2013), but less consistently in adults (Schendan et al., 1998; Xue et al., 2008). The current study takes these previous findings a step further by investigating print tuning in children and adults within the same paradigm and experimental task demands.

In accordance with the literature (e.g., Brem et al., 2013; Maurer et al., 2006), robust print tuning that occurred in the early and late N1 segments was observed in primary school children. While print tuning was more pronounced in the late than in the early N1 in children, it was present in the early N1 in adult participants. This shift of print tuning from predominantly late N1 in children to the early N1 in adults could be indicative of either a developmental change in the dynamics of visual word processing or of structural and morphologic changes of the brain related to maturation (see S3). While the largest changes in the brain volume happen in the early childhood (Giedd, 2004), these maturational changes continue until early adulthood and include alterations in gray and white matter (Giedd et al., 1999; Giedd, 2004; Paus et al, 1999; Sowell et al., 2001; 2004), brain growth in language-related areas, and ameliorations in connectivity due to synaptic pruning (Sowell et al., 2001). However, given that maturational latency-decreases should affect the speed of both word and false-font processing to a similar extent, the observed latency shift could reflect the increase in reading skills from childhood to adulthood.

The two-phasic N1 in adults with larger amplitudes to German words than false-font strings in the early segment and the lack of such sensitivity in the late segment (possibly due to an increased processing speed for words) may reflect the influence of expertise on visual processing. Through extensive reading experience, adult readers develop visual expertise for words, which presumably leads to faster recognition of familiar visual patterns and thus to enhanced early activation of words during the N1 component. This is in agreement with previous studies showing an N1 latency difference between words and unfamiliar visual control stimuli (Brem et al., 2006; Maurer et al., 2008; Shirahama et al., 2004). Although the waveforms in our study seemed to indicate a similar latency difference between German words and false-font strings, a peak latency analysis showed no significant difference (see S4). This suggests that differences in processing speed do not necessarily need to be reflected in the peak latency of a given component. Given that visual word patterns may be processed early in the course of N1, in a later phase, other word properties (such as lexical validity, discussed below) could come into play. On the other hand, for the visually more complex false-font characters, the later N1 phase would correspond to more basic visual processes. Thus, the statistically not significant word–false-font difference in the late N1 over the left hemisphere and at the same time a reversed effect over the right hemisphere (as indicated by the t-maps) may also reflect higher visual processing demands that are necessary to detect repetitions of the visually more complex false-font strings (compared to symbol strings in previous studies). This may explain some of the inconsistent results on print tuning when closely matched visual control stimuli were used (Eulitz et al., 2000; Schendan et al., 1998; Wong et al., 2005; Xue et al., 2008). However, despite the well-matched control condition in the present study, print tuning could be reliably detected in the early N1 segment in adults. Previous studies showed that in contrast to adults, print tuning was robustly found in children, irrespective of the use of simple geometric forms (e.g., Araújo et al, 2012; Maurer et al., 2006; Parviainen et al., 2006) or false-font strings (Brem et al., 2010; Brem et al., 2013; Hasko et al., 2013). As in the present study the false-font N1 was of rather similar amplitude in children and adults (Fig. 2), the robust print tuning effect in children was due to the increased N1 in the

German word condition. This is in agreement with the development of perceptual expertise in novice and expert learners (Palmeri, Wong, & Gauthier, 2004) and with the finding of an inverted u-curve of the developmental course of print tuning (Brem et al., 2009; Maurer et al., 2006).

4.2. Lexicality effects – found only in adults' native language

The current study indicated that adult readers process lexical word properties in the late N1 segment. However, this was true for the native but not for the second language, which concurs with a study by Proverbio et al. (2009), who reported N1 lexicality effects to be more pronounced in the native than non-native languages. Notably, the direction of the L1 lexicality effect, i.e. larger N1 for German words than pseudowords, was as expected based on previous studies using a task that involved reading only implicitly (Maurer et al., 2005b; Proverbio et al., 2009). In the context of two neural models of orthographic processing i.e. "Local Combination Detector Model" (LCD, Dehaene et al., 2005; Dehaene, & Cohen, 2011) and "Interactive Account Model" (Price & Devlin, 2011), the direction of this effect could only be explained by task effects. According to the LCD Model (Dehaene et al., 2005; Dehaene, & Cohen, 2011) bottom-up tuning to orthographic features would be expected to be similar for matched words and pseudowords. According to the Interactive Account (Price & Devlin, 2011) the activation to pseudowords would be expected to be larger than activation for words due to more prediction errors from higher-level language regions to pseudowords. Thus, in the context of an implicit reading task, words might engage automatic reading-related processes to a larger degree than pseudowords do (Maurer et al., 2005b). Such a task-related interpretation would be in agreement with findings of reversed lexical effects in explicit reading tasks, such as lexical decision (Hauk et al., 2006b; Hauk et al., 2006a). Similar reversed effects of lexical familiarity were found in studies comparing low- and high-frequency words (Assadollahi & Pulvermuller, 2003; Hauk & Pulvermuller, 2004, Rugg, 1990; Sereno et al., 1998). Larger activation for pseudowords than words, as well as for low-frequency than high-frequency words in these studies could be explained in terms of strategic top-down modulations induced by task demands requiring thorough examination of pseudowords or less familiar words (Price et al., 2011). On the other hand, there are lexical decision studies showing the reversed pattern in the late N1 (i.e. larger activation for words than pseudowords; Mahé et al., 2012; 2013; Rosazza et al., 2009). As the lexical decision tasks in these studies did not only include words and pseudowords, but also consonant strings (Mahé et al., 2012, 2013; Rosazza et al. 2009) and symbols (Mahé et al., 2012, 2013), lexicality effects in the N1 may not only be influenced by task demands, but also by the context. Hence, interactions between task and context may potentially explain some of the null effects for word-pseudoword differences in several previous studies (Araújo et al., 2012; Bentin et al., 1999; Kast et al., 2010; Maurer et al., 2005a) and shall thus be addressed in future investigations.

In contrast to a previous study that applied the same task and a similar paradigm (Maurer et al., 2006), and contrary to our expectation, no lexicality effects were found in children irrespective of early or late N1 segment. This was true for both the native and the second language. Similarly, the t-maps of the lexicality contrast in children showed that the electrodes with the highest t-values did not correspond to the typical N1 distribution, unlike in adults' late N1, where the locations were in occipito-temporal and fronto-central regions.

In contrast to adults, where words may engage automatic language-related processes to a larger degree in an implicit reading task than pseudowords do, this might not yet be the case in children. Such language-related processes may require more time to be established and only become fully automatized through intensive reading practice over several years. The lack of such automaticity may render the N1 lexicality effects more susceptible to interference through other factors, such as attention or fatigue. This could also explain why N1 lexicality effects were not significant in previous studies with children (Araújo et al., 2012; Kast et al., 2010). Moreover, due to the use of a simple repetition detection task we cannot rule out the possibility that the children tested perceived the pronounceable pseudowords as actual words that were not yet incorporated into their mental lexicon. Even though another previous study using a lexical decision task reported a similar lack of lexicality effects in children aged 8 to 12 (Kast et al., 2010), it is plausible that a lexical decision task would be more appropriate to test such young children. Future studies shall thus clarify whether different tasks would yield N1 differences in processing of lexical properties. Notably, due to the low number of English words (see section 2.3.) each stimulus was repeated six times, which might have reduced the novelty of pseudowords (possibly making it act like low-frequency words at the orthographic and phonological level) and thus attenuated the size of N1 lexicality effects. However, this notion was not supported by our data, as an analysis comparing L1 and L2 lexicality effects (for details see S5) did not reveal significant effects of repetition neither in children nor in adults (all $p \geq .100$, see also S5).

4.3. Limitations

While it has been shown that sublexical characteristics like syllable-frequency as well as orthographic and phonological neighbourhood may affect the time course of print processing (e.g., Barber et al., 2004; Bürki et al., 2015; Chetail et al., 2012; Hutzler et al., 2004), being restricted to a very small number of English words (see also Methods), we could not have controlled for all these potentially important linguistic factors of the stimuli. Hence, while studying early print processing we found it of utmost importance to control for lexical familiarity and visual complexity of the stimuli presented. Furthermore, it is worth noting that there are different approaches to define time windows of interest. In the present study the N1 component was subdivided into its early and late parts based on previous studies (Appelbaum et al., 2009; Cohen et al., 2000; Korinth, et al., 2011, 2012; Maurer et al., 2005a; Nemrodov, et al., 2011; Simon et al., 2004). While this has been done by dividing the N1 time window into two segments of equal length (see Methods), it is possible that other approaches that take the topographic distributions into account might reveal more precise boarders between the early and late part of the N1. Similarly, based on previous studies (e.g., Bentin et al., 1999; Brem et al., 2005; Eberhard-Moscicka et al., 2015) our analysis focused on effects that corresponded to a typical N1 topography with occipito-temporal negativity (and fronto-central positivity). A topographic whole map approach might reveal additional effects that do not correspond to the N1 topography, but would also bear the risk of missing consistent, but focal effects, as the occipito-temporal electrodes were more sensitive to N1 effects than global measures in previous studies (Maurer et al., 2007; 2011).

5. Conclusions

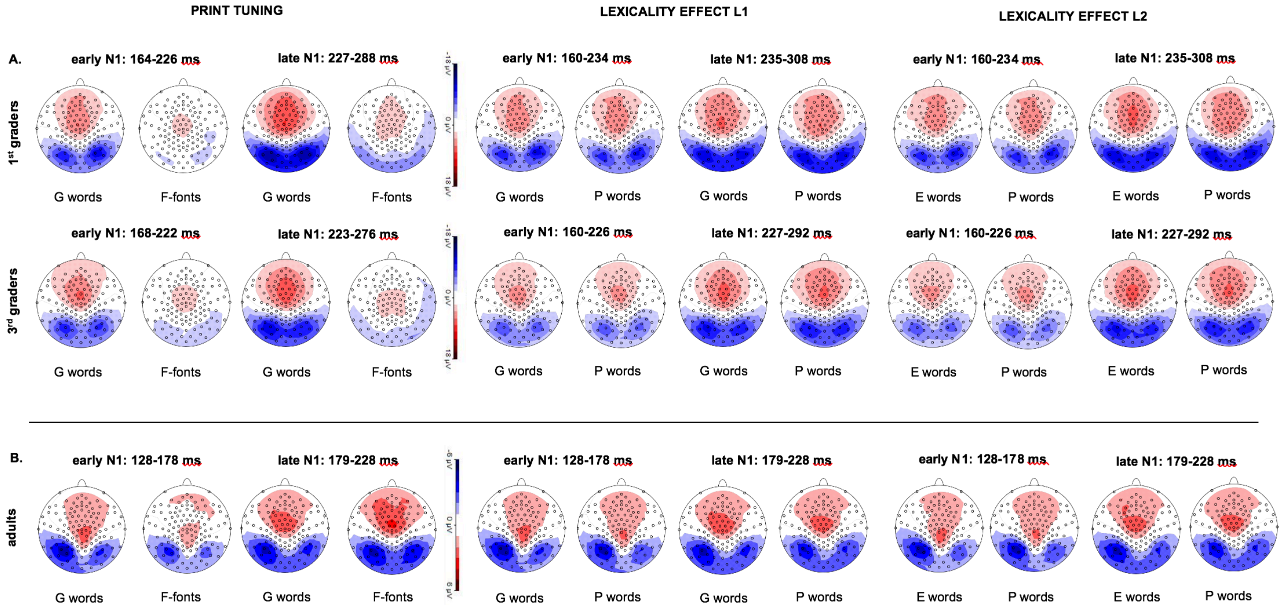
The longitudinal data of children indicated robust print tuning effects across both early and late N1 segments, but no lexicality effects in neither early nor late N1 segment. While adult speakers of English as a foreign language exhibited print tuning in the early N1, lexicality effects reached significance in the late N1. However, the lexicality effects were found only for the native but not for the L2, confirming previous results showing stronger lexicality effects for the native language compared to the L2 (Proverbio et al., 2009). The different temporal pattern of the adults' print tuning and lexicality effects suggests that in expert readers different aspects of orthographic processing converge within the time range of the N1 component with more basic visual aspects being processed early and more high-level language aspects being processed late. Such dynamic early visual word processing of expert readers may take years to develop in children learning to read or in adults learning a foreign language.

6. Supporting information

S1. List of stimuli presented in the one-back task.

German words	False-font strings	Pseudowords	English words
BEIN	ΩΔ[Λ	BREET	CHILD
KIND	Ц[ΛΒ	CHARN	MOON
WORT	+⊡ƚℸ	HOID	DESK
MANN	⊬⊡ΛΛ	WOON	BAG
KNIE	ЦΛ[Δ	GED	KING
RAD	ƚ⊡Β	MINS	DOLL
MILCH	⊬[Λ>Ɣ	RAS	TREE
PFERD	♆◇ΔƚΒ	TELN	CAT
OHR	⊡Ɣƚ	DAP	GIRL
BOOT	Ω⊡⊡ℸ	KILG	PLANT
SPIEL	Ω♆[ΔΛ	QUOP	DOOR
GELD	◇ΔΛΒ	GOND	SUN
HUT	ƔΠℸ	TID	QUEEN
TAG	ℸ⊡◇	KRELL	DOG

The words Kind, Boot, Hut and Tag are valid in German and in English. To verify that these were perceived as German words, six independent participants were asked to read out loud all the word stimuli from the one-back task (in the exact same experiment as in the EEG session). One out of 6 participants read the word “Boot” according to English pronunciation (in 5 out of 6 presentations). This means that across all 6 participants and all presentations, less than 1% of all German words presented were read according to English pronunciation (i.e., over 99% were read as German words). Notably, this participant studies informatics where the English word “boot” belongs to professional vocabulary (occurring in German texts with an English pronunciation). Hence, the participants’ native German language together with the blocked stimulus presentation induced a strong bias towards German pronunciation of the ambiguous words in the German condition.



S2. Early and late N1 segment maps for German words (G words), false-font strings (F-fonts), English words (E words) and pseudowords (P words) illustrated separately for print tuning and lexicality effects in L1 (native language) and L2 (second language) for 1st and 3rd graders (A) and for adults (B).

S3. Analysis comparing 3rd grade children to adults.
In a supplementary analysis we directly compared temporal dynamics of print tuning and lexicality effects within the N1 time range at left occipito-temporal electrodes between 3rd grade children and adults by computing a repeated-measure ANOVA with the within-subject factors *effect* (print tuning vs. lexicality L1) and *segment* (early N1 vs. late N1), and a between-subject factor *group* (3rd grade vs. adults). The results revealed the critical three-way interaction (*segment* x *effect* x *group*, $F(1,47)=67.74$, $p<.001$, $\eta_p^2=.58$). This interaction indicated that the differential temporal dynamics of the print tuning and lexicality effects across early and late N1 segments in adults did not yet occur in children. This three-way interaction further modulated the main effects of *group* ($F(1,47)=21.09$, $p<.001$, $\eta_p^2=.31$), *segment* ($F(1,47)=12.56$, $p=.001$, $\eta_p^2=.21$) and *effect* ($F(1,47)=78.77$, $p<.001$, $\eta_p^2=.63$), as well as the two-way interactions *segment-by-effect* ($F(1,47)=10.09$, $p=.003$, $\eta_p^2=.18$), *segment*-by-*group* ($F(1,47)=33.50$, $p<.001$, $\eta_p^2=.42$), and *effect*-by-*group* ($F(1,47)=68.28$, $p<.001$, $\eta_p^2=.59$).

S4. Supplementary peak latency analysis.
To investigate differences in peak latencies between German words and false-font strings, German words and pseudowords as well as English words and pseudowords in adult participants, we performed automatic peak detection (with the use of Brain Vision Analyzer Software) for the N1 time segment of interest (128-228 ms). Further, we computed separate ANOVAs on the left occipito-temporal cluster for the latency in print tuning (*wordlike*: German words vs. false-font strings; $F(1,21)=2.40$, $p=.136$), latency in the lexicality effect L1 (*lexicality L1*: German words vs. pseudowords; $F(1,21)=0.00$, $p=.955$) as well as latency in the lexicality effect L2 (*lexicality L2*: English words vs. pseudowords; $F(1,21)=1.96$, $p=.663$). These results indicated no significant latency differences for neither of the effects investigated.

S5. Supplementary analysis of the repetition effect.
To investigate the effect of repetition, L1 and L2 lexicality effects were compared in a three-way ANOVA with factors: *effect* (lexicality L1 vs. lexicality L2), *repetition* (ERPs of the 1st half of the EEG vs. ERPs of the 2nd half of the EEG) and *segment* (N1 early vs. N1 late), separately for each group (i.e. adults, 1st and 3rd graders). These ANOVAs did not reveal significant effects of repetition neither in children nor in adults (all $p\geq.100$).

Acknowledgements
This research was supported by the Swiss National Science Foundation (SNF-128610). We would like to thank the anonymous reviewers for their valuable input.

References

Aghababian, V., Nazir, T.A., 2000. Developing normal reading skills: aspects of the visual processes underlying word recognition. J. Exp. Child Psychol. 76 (2), 123–150.

Albrecht, B., Banaschewski, T., Brandeis, D., Heinrich, H., Rothenberger, A., 2005. Response inhibition deficits in externalizing child psychiatric disorders: an ERP-study with the stop-task. Behav. Brain Funct. 1 (1), 1–22.

Appelbaum, L.G., Liotti, M., Perez, R., Fox, S.P., Woldorff, M.G., 2009. The temporal dynamics of implicit processing of non-letter, letter, and word-forms in the human visual cortex. Front. Hum. Neurosci., 3.

Araújo, S., Bramão, I., Faísca, L., Petersson, K.M., Reis, A., 2012. Electrophysiological correlates of impaired reading in dyslexic pre-adolescent children. Brain Cogn. 79 (2), 79–88.

Assadollahi, R., Pulvermuller, F., 2003. Early influences of word length and frequency: a group study using MEG. NeuroReport 14 (8), 1183–1187.

Baayen, R., Piepenbrock, R., Van Rijn, H., 1993. The CELEX Lexical Database [CD- ROM]. University of Pennsylvania Linguistic Data Consortium, Philadelphia.

Baker, C.I., Liu, J., Wald, L.L., Kwong, K.K., Benner, T., Kanwisher, N., 2007. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. Proc. Natl. Acad. Sci. USA 104 (21), 9087–9092.

Barber, H., Vergara, M., Carreiras, M., 2004. Syllable-frequency effects in visual word recognition: evidence from ERPs. NeuroReport 15 (3), 545–548.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M.H., Echallier, J.F., Pernier, J., 1999. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. J. Cogn. Neurosci. 11 (3), 235–260.

Brem, S., Lang-Dullenkopf, A., Maurer, U., Halder, P., Bucher, K., Brandeis, D., 2005. Neurophysiological signs of rapidly emerging visual expertise for symbol strings. Neuroreport 16 (1), 45–48.

Brem, S., Bucher, K., Halder, P., Summers, P., Dietrich, T., Martin, E., et al., 2006. Evidence for developmental changes in the visual word processing network beyond adolescence. Neuroimage 29 (3), 822–837.

Brem, S., Halder, P., Bucher, K., Summers, P., Martin, E., Brandeis, D., 2009. Tuning of the visual word processing system: distinct developmental ERP and fMRI effects. Hum. Brain Mapp. 30 (6), 1833–1844.

Brem, S., Bach, S., Kucian, K., Guttorm, T.K., Martin, E., Lyytinen, H., et al., 2010. Brain sensitivity to print emerges when children learn letter–speech sound correspondences. Proc. Natl. Acad. Sci. USA 107 (17), 7939–7944.

Brem, S., Bach, S., Kujala, J.V., Maurer, U., Lyytinen, H., Richardson, U., et al., 2013. An electrophysiological study of print processing in kindergarten: the contribution of the visual N1 as a predictor of reading outcome. Dev. Neuropsychol. 38 (8), 567–594.

Bürki, A., Cheneval, P.P., Laganaro, M., 2015. Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. Brain Lang. 150, 90–102.

Chetail, F., Colin, C., Content, A., 2012. Electrophysiological markers of syllable frequency during written word recognition in French. Neuropsychologia 50 (14), 3429–3439.

Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. Neuroimage 22 (1), 466–476.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., et al., 2000. The visual word form area: spatial and temporal characteriza- tion of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123 (Pt 2), 291–307.

Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn. Sci. 15 (6), 254–262.

Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. Trends Cogn. Sci. 9 (7), 335–341.

Duyck, W., Desmet, T., Verbeke, L.P.C., Brysbaert, M., 2004. WordGen: a tool for word selection and nonword generation in Dutch, English, German, and French. Behav. Res. Methods Instrum. Comput. 36, 488–499.

Eberhard-Moscicka, A.K., Jost, L.B., Raith, M., Maurer, U., 2015. Neurocognitive me- chanisms of learning to read: print tuning in beginning readers related to word- reading fluency and semantics but not phonology. Dev. Sci. 18 (1), 106–118.

Eulitz, C., Eulitz, H., Maess, B., Cohen, R., Pantev, C., Elbert, T., 2000. Magnetic brain activity evoked and induced by visually presented words and nonverbal sti- muli. Psychophysiology 37 (4), 447–455.
Ferree, T.C., Luu, P., Russell, G.S., Tucker, D.M., 2001. Scalp electrode impedance, infection risk, and EEG data quality. Clin. Neurophysiol. 112 (3), 536–544.
Franklin, M.S., Dien, J., Neely, J.H., Huber, E., Waterson, L.D., 2007. Semantic priming modulates the N400, N300, and N400RP. Clin. Neurophysiol. 118 (5), 1053–1068.
Giedd, J.N., 2004. Structural magnetic resonance imaging of the adolescent brain. Ann. N. Y. Acad. Sci. 1021 (1), 77–85.
Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Ra- poport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nat. Neurosci. 2 (10), 861–863.
Gros, H., Doyon, B., Rioual, K., Celsis, P., 2002. Automatic grapheme processing in the left occipitotemporal cortex. Neuroreport 13 (8), 1021–1024.
Grossi, G., Savill, N., Thomas, E., Thierry, G., 2010. Posterior N1 asymmetry to English and Welsh words in Early and Late English–Welsh bilinguals. Biol. Psychol. 85 (1), 124–133.
Hämäläinen, J.A., Lohvansuu, K., Ervast, L., Leppänen, P.H., 2015. Event-related po- tentials to tones show differences between children with multiple risk factors for dyslexia and control children before the onset of formal reading instruction. Int. J. Psychophysiol. 95 (2), 101–112.
Hasko, S., Groth, K., Bruder, J., Bartling, J., Schulte-Körne, G., 2013. The time course of reading processes in children with and without dyslexia: an ERP study. Front. Hum. Neurosci., 7.
Hauk, O., Pulvermuller, F., 2004. Effects of word length and frequency on the human event-related potential. Clin. Neurophysiol. 115 (5), 1090–1103. http://dx.doi. org/10.1016/j.clinph. 2003.12.020.
Hauk, O., Davis, M.H., Ford, M., Pulvermuller, F., Marslen-Wilson, W.D., 2006a. The time course of visual word recognition as revealed by linear regression analysis of ERP data. Neuroimage 30 (4), 1383–1400.
Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermuller, F., Rogers, T.T., 2006b. [Q] When would you prefer a SOSSAGE to a SAUSAGE? [A] At about 100 ms. ERP correlates of orthographic typicality and lexicality in written word recognition. J. Cogn. Neurosci. 18 (5), 818–832.

Hutzler, F., Bergmann, J., Conrad, M., Kronbichler, M., Stenneken, P., Jacobs, A.M., 2004. Inhibitory effects of first syllable-frequency in lexical decision: an event- related potential study. Neurosci. Lett. 372 (3), 179–184.
Jost, L.B., Eberhard-Moscicka, A.K., Pleisch, G., Heusser, V., Brandeis, D., Zevin, J., et al., 2015. Native and non-native speech sound processing and the neural mismatch responses: a longitudinal study on classroom-based foreign language learning. Neuropsychologia 72, 94–104.
Kast, M., Elmer, S., Jancke, L., Meyer, M., 2010. ERP differences of pre-lexical pro- cessing between dyslexic and non-dyslexic children. Int. J. Psychophysiol. 77 (1), 59–69.
Korinth, S.P., Sommer, W., Breznitz, Z., 2011. Toward an ERP-driven diagnostic ap- proach for reading impairments. Dev. Neuropsychol. 36 (7), 944–948.
Korinth, S.P., Sommer, W., Breznitz, Z., 2012. Does silent reading speed in normal adult readers depend on early visual processes? Evidence from event-related brain potentials. Brain Lang. 120 (1), 15–26.
Laufer, B., Nation, P., 1999. A vocabulary-size test of controlled productive ability. Lang. Test. 16 (1), 33–51.
Lehmann, D., Skrandies, W., 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr. Clin. Neurophysiol. 48 (6), 609–621.
Luu, P., Ferree, T., 2005. Determination of the HydroCel Geodesic Sensor Nets' Average Electrode Positions and Their 10–10 International Equivalents.Inc, Technical Note.
Mahé, G., Bonnefond, A., Doignon-Camus, N., 2013. Is the impaired N170 print tuning specific to developmental dyslexia? A matched reading-level study with poor readers and dyslexics. Brain Lang. 127 (3), 539–544.
Mahé, G., Bonnefond, A., Gavens, N., Dufour, A., Doignon-Camus, N., 2012. Impaired visual expertise for print in French adults with dyslexia as shown by N170 tuning. Neuropsychologia 50 (14), 3200–3206.
Maurer, U., McCandliss, B.D., 2007. The development of visual expertise for words: the contribution of electrophysiology. In: Grigorenko, E.L., Naples, A.J. (Eds.), Single-Word Reading: Biological and Behavioral Perspectives. Lawrence Erl- baum Associates, Mahwah, NJ.

Maurer, U., Brandeis, D., McCandliss, B.D., 2005a. Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. Behav. Brain Funct. 1 (1), 13.
Maurer, U., Zevin, J.D., McCandliss, B.D., 2008. Left-lateralized N170 effects of visual expertise in reading: evidence from Japanese syllabic and logographic scripts. J. Cogn. Neurosci. 20 (10), 1878–1891.
Maurer, U., Brem, S., Bucher, K., Brandeis, D., 2005b. Emerging neurophysiological specialization for letter strings. J. Cogn. Neurosci. 17 (10), 1532–1552.
Maurer, U., Blau, V.C., Yoncheva, Y.N., McCandliss, B.D., 2010. Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. Dev. Neuropsychol. 35 (4), 404–422.
Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., et al., 2006. Coarse neural tuning for print peaks when children learn to read. Neuroimage 33 (2), 749–758.
Maurer, U., Schulz, E., Brem, S., der Mark, S.V., Bucher, K., Martin, E., et al., 2011. The development of print tuning in children with dyslexia: Evidence from long- itudinal ERP data supported by fMRI. Neuroimage 57 (3), 714–722.
Mayringer, H., Wimmer, H., 2005. Salzburger Lese-Screening für Die Klassenstufen 1-4. Manual. Hans Huber, Bern.
McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. 7 (7), 293–299.
McLaughlin, J., Osterhout, L., Kim, A., 2004. Neural correlates of second-language word learning: minimal instruction produces rapid change. Nat. Neurosci. 7 (7), 703–704.
Meyer, M., Elmer, S., Baumann, S., Jancke, L., 2007. Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. Restor. Neurol. Neurosci. 25 (3–4), 411–431.
Moll, K., Landerl, K., 2010. SLRT-II Lese- und Rechtschreibtest. Hans Huber, Bern.
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385 (6615), 432–434.
Nemrodov, D., Harpaz, Y., Javitt, D.C., Lavidor, M., 2011. ERP evidence of hemispheric independence in visual word recognition. Brain Lang. 118 (3), 72–80.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al., 2008. Second-language learning and changes in the brain. J. Neurolinguist. 21 (6), 509–521.
Palmeri, T.J., Wong, A.C., Gauthier, I., 2004. Computational approaches to the development of perceptual expertise. Trends Cogn. Sci. 8 (8), 378–386.
Parviainen, T., Helenius, P., Poskiparta, E., Niemi, P., Salmelin, R., 2006. Cortical sequence of word perception in beginning readers. J. Neurosci. 26 (22), 6052–6061.
Paus, T., Zijdenbos, A., Worsley, K., Collins, D.L., Blumenthal, J., Giedd, J.N., Evans, A. C., 1999. Structural maturation of neural pathways in children and adolescents: in vivo study. Science 283 (5409), 1908–1911.
Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., Dehaene, S., 2014. Timing the impact of literacy on visual processing. Proc. Natl. Acad. Sci. USA 111, 49.
Petermann, F., Petermann, U., 2010. HAWIK-IV Hamburg-Wechsler-Intelligenztest für Kinder – IV. Hans Huber, Bern.
Price, C.J., Devlin, J.T., 2011. The interactive account of ventral occipitotemporal contributions to reading. Trends Cogn. Sci. 15 (6), 246–253.
Proverbio, A.M., Čok, B., Zani, A., 2002. Electrophysiological measures of language processing in bilinguals. J. Cogn. Neurosci. 14 (7), 994–1017.
Proverbio, A.M., Adorni, R., Zani, A., 2009. Inferring native language from early bio- electrical activity. Biol. Psychol. 80 (1), 52–63.
Rihs, T.A., Michel, C.M., Thut, G., 2007. Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. Eur. J. Neurosci. 25 (2), 603–610.
Rinker, T., Alku, P., Brosch, S., Kiefer, M., 2010. Discrimination of native and non- native vowel contrasts in bilingual Turkish–German and monolingual German children: insight from the mismatch negativity ERP component. Brain Lang. 113 (2), 90–95.
Rosazza, C., Cai, Q., Minati, L., Paulignan, Y., Nazir, T.A., 2009. Early involvement of dorsal and ventral pathways in visual word recognition: an ERP study. Brain Res. 1272, 32–44.
Rossion, B., Joyce, C.A., Cottrell, G.W., Tarr, M.J., 2003. Early lateralization and or- ientation tuning for face, word, and object processing in the visual cortex. Neuroimage 20, 1609–1624.
Rugg, M.D., 1990. Event-related brain potentials dissociate repetition effects of high-and low-frequency words. Mem. Cogn. 18 (4), 367–379.

Schendan, H.E., Ganis, G., Kutas, M., 1998. Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. Psychophysiology 35 (3), 240–251.
Schroeder, S., Würzner, K.M., Heister, J., Geyken, A., Kliegl, R., 2015. ChildLex: A lexical database of German read by children. Behavior research methods 47 (4), 1085–1094.
Sereno, S.C., Rayner, K., Posner, M.I., 1998. Establishing a time-line of word re- cognition: evidence from eye movements and event-related potentials. Neu- roReport 9 (10), 2195–2200.
Shestakova, A., Huotilainen, M., Ceponiene, R., Cheour, M., 2003. Event-related potentials associated with second language learning in children. Clin. Neuro- physiol. 114 (8), 1507–1512.
Shirahama, Y., Ohta, K., Takashima, A., Matsushima, E., Okubo, Y., 2004. Magnetic brain activity elicited by visually presented symbols and Japanese characters. Neuroreport 15 (5), 771–775.
Simon, G., Bernard, C., Largy, P., Lalonde, R., Rebai, M., 2004. Chronometry of visual word recognition during passive and lexical decision tasks: an ERP investiga- tion. Int. J. Neurosci. 114 (11), 1401–1432.
Sowell, E.R., Thompson, P.M., Tessner, K.D., Toga, A.W., 2001. Mapping continued brain growth and gray matter density reduction in dorsal frontal cortex: in- verse relationships during postadolescent brain maturation. J. Neurosci. 21 (22), 8819–8829.
Sowell, E.R., Thompson, P.M., Leonard, C.M., Welcome, S.E., Kan, E., Toga, A.W., 2004. Longitudinal mapping of cortical thickness and brain growth in normal chil- dren. J. Neurosci. 24 (38), 8223–8231.
Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., Koenig, T., 2006. Plasticity in the adult language system: a longitudinal electrophysiological study on second language learning. Neuroimage 33 (2), 774–783.
Strimmer, K., 2008. A unified approach to false discovery rate estimation. BMC Bioinform. 9 (1), 303.
Tarkiainen, A., Helenius, P., Hansen, P.C., Cornelissen, P.L., Salmelin, R., 1999. Dy- namics of letter string perception in the human occipitotemporal cortex. Brain 122 (11), 2119–2132.
Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al., 1999. Brain responses reveal the learning of foreign language phonemes. Psycho- physiology 36 (5), 638–642.
Wong, A.C., Gauthier, I., Woroch, B., DeBuse, C., Curran, T., 2005. An early electro- physiological response associated with expertise in letter perception. Cogn. Affect. Behav. Neurosci. 5 (3), 306–318.

Xue, G., Jiang, T., Chen, C., Dong, Q., 2008. Language experience shapes early elec- trophysiological responses to visual stimuli: the effects of writing system, sti- mulus length, and presentation duration. Neuroimage 39 (4), 2025–2037.

Yoncheva, Y.N., Maurer, U., Zevin, J.D., McCandliss, B.D., 2013. Effects of rhyme and spelling patterns on auditory word ERPs depend on selective attention to phonology. Brain Lang. 124 (3), 238–243.

Zhao, J., Kipp, K., Gaspar, C., Maurer, U., Weng, X., Mecklinger, A., et al., 2014. Fine neural tuning for orthographic properties of words emerges early in children reading alphabetic script. J. Cogn. Neurosci. 26 (11), 2431–2442.